

М. Ю. МУХИН

*(Уральский федеральный университет им. первого Президента России Б. Н. Ельцина,
г. Екатеринбург, Россия)*

УДК 811.161.1'42:81'23:821161.1-3

НА ПЕРЕКРЕСТКЕ СТАТИСТИКИ, ИДИОСТИЛИСТИКИ И ПСИХОЛИНГВИСТИКИ²⁰

Аннотация. В статье обзревается некоторые возможности использования статистических данных в лингвистике текста и идиостилистике. На основании интерпретации этих данных возможны психолингвистические выводы, касающиеся авторов художественных текстов. Обсуждаются признаки текстов, особенно формальные, которые наиболее пригодны для статистических измерений. Подробно, не только с количественной, но и с качественной точки зрения, рассмотрен параметр объема художественного произведения (на материале русской прозы XIX-XX вв.). На материале прозы XX в. (В. Набоков, А. Платонов, М. Булгаков, М. Шолохов) анализируется распределение слов разных частей речи, средняя длина предложения и условный словарный запас — коэффициент типичности знаков — в произведениях разных авторов. Представлены сведения о различиях в статистических параметрах, которые свидетельствуют об идиостилевых и психолингвистических установках русских прозаиков XIX-XX веков.

Ключевые слова: статистический анализ, идиостиль, лингвистический анализ текста, психолингвистика, В. Набоков, А. Платонов, М. Булгаков, М. Шолохов, русская литература.

Элементы статистики давно и вполне эффективно используются в филологических исследованиях. Без них не обходится психо- и социолингвистические эксперименты, а объективировать работы, выполненные в русле ассоциативной семантики (в особенности, фоносемантики), без статистических данных практически невозможно. Статистические методы «алгеброй» про-

²⁰ Исследование выполнено при поддержке Программы повышения конкурентоспособности Уральского федерального университета (номер соглашения 02.А03.21.0006).

ники и в классическую филологию, которая изучает то, что со времен Пушкина мы называем «гармонией» — художественную литературу. Все счетные и дискретные единицы языка, употребляемые в тексте, могут быть элементами статистического анализа: различные грамматические и лексические аспекты, формальные составляющие стихотворного ритма и рифмы. Масштабные исследования измеряют динамику развития индивидуального стиля и литературных направлений, структуру целой литературной ситуации — см., например, работы [Buttows 1992; Баевский 2001] и др.

Однако требования статистики предполагают строгую формализацию текстовых единиц, что затрудняет использование статистических методов в плане интерпретации смысла текста. Безусловно, лингвистическая теория развивается в направлении все большей системности текстового анализа, чему способствует и распространение влияния корпусной лингвистики. Различные системные модели анализа текста представлены в работах [Бабенко 2004; Болотнова 1992; Казарин 2002; Лукин 1999; Новиков 2007; Пищальникова 1992; Сулименко 2009] и многих др. Следует отметить, что представленные в этих монографиях и учебных пособиях структурные модели анализа текста не являются настолько строгими, чтобы их можно было использовать в целях автоматической обработки текста, который бы становился вследствие применения этих моделей машиночитаемым. Иначе говоря, филологические модели все равно ориентированы на использование их человеком — тем более с гуманитарным опытом текстовой интерпретации. В самом деле, моделируя макрокатегории текста (образы автора и персонажей, хронотоп и т. д.), мы обязательно предполагаем человеческое прочтение художественного текста. Ведь восприятие литературного произведения, помимо оценки его формы и содержания, предполагает *переживание* текста читателем. Между делом эта эмоциональная составляющая проникает в коммуникативные модели текста — см., к примеру, работы И. Я. Чернухиной, в которых предложен термин «эффект неожиданности», возникающий вследствие ложного читательского прогнозирования [Чернухина 1990: 174].

Художественную литературу как никакой другой объект лингвистики трудно квантифицировать в силу наличия в нем особо-

го образования, который М. М. Бахтин называл «эстетическим объектом». Однако и художественный текст характеризуется измеряемостью. Он состоит из языковых единиц, имеющих количественные и качественные признаки, причем первые, естественно, формализуются значительно проще. От них мы и будем отталкиваться в рассуждении о нашем предмете, который находится «на перекрестке статистики, идиостилистики и психолингвистики». Наша задача – обсудить ряд статистических параметров текста, позволяющих судить не только об особенностях идиостиля, но и, возможно, о психологических особенностях автора – по крайней мере того условного автора, который представлен в литературных произведениях.

Рассмотрим наиболее простые для статистических измерений параметры художественных текстов:

- 1) объем произведения,
- 2) распределение слов разных частей речи и средняя длина предложения,
- 3) лексика, авторский словарный запас.

1. Объем произведения

В Интернете можно обнаружить разные версии топ-списка самых крупных художественных текстов. По одной из них, самым длинным произведением в мире является роман Жюль Ромена «Люди доброй воли» («Les Hommes de bonne volonté»), написанный с 1932 по 1946 гг. Роман состоит из 27 томов и, кроме того, содержит огромный страничный указатель. По некоторым данным, объем этого произведения превышает два миллиона слов, т.е. он по объему больше, чем увеличенный в четыре раза роман Л. Н. Толстого «Война и мир».

С точки зрения классической филологии объем текста – достаточно факультативная характеристика, по крайней мере если мы берем один жанр (например, сопоставляем между собой только романы или рассказы); главное для филолога – взгляд на текст с точки зрения поэтики. Однако, по нашему мнению, ценностных признаков у объема художественного произведения гораздо больше, чем кажется на первый взгляд.

Во-первых, для статистических измерений: очевидно, что чем больше по объему текст, тем результаты исследования объ-

ективнее, так как повторяемость единиц выше. В этом смысле текстовые элементы можно расположить по шкале частоты встречаемости. Чем частотнее явление, тем меньше погрешность при интерпретации статистических данных. Самые частотные – буквы и буквосочетания; затем – морфологические признаки слов, синтаксические конструкции и связи; далее – лексемы (сначала самые частотные: служебные слова, местоимения, – потом знаменательная лексика от глагола *быть*, по мере убывания частоты). И замыкают список – элементы композиции, которых по определению не может быть много, даже в большом романе.

Во-вторых, с объемом связана вероятность стилизации, подделки. Роман большого объема в духе того или иного автора написать значительно сложнее, чем небольшое стихотворение. В третьих, объем произведения можно считать идиостилевым показателем в пиковых случаях (примеры из русской литературы – Л. Н. Толстой и М. А. Шолохов) или в случае повторяющихся форм, какой является, например, онегинская строфа.

Посмотрим на вычисленные нами примерные объемы классической прозы XIX века.

Л. Н. Толстой: «Война и Мир» (около 459 тыс. слов), «Анна Каренина (около 270), «Воскресение» (136,5). Ф. М. Достоевский: Братья Карамазовы (290,5), «Игрок» (225), «Бесы» (198), «Подросток» (187), «Преступление и наказание» (184). И. А. Гончаров: «Фрегат “Паллада”» (238,4), «Обрыв» (232,5), «Обломов» (150), «Обыкновенная история» (94).

И на фоне перечисленных авторов — И. С. Тургенев: «Отцы и дети» (55,5), «Дворянское гнездо» (47,4), «Рудин» (36,2). Н. В. Гоголь: «Мертвые души» (72,8), М. Ю. Лермонтов: «Герой нашего времени» (41,8).

Сравнивать по объему произведения разных веков некорректно: в XX в. стали писать в среднем значительно менее крупные романы (см. далее), но для одного века цифры вполне показательны. Писатели, склонные к (условно) более эпичному повествованию, включающему большое количество фактов, персонажей, событий, создают и более объемные тексты. И в этом, безусловно, проявляются личностные черты «эпиков» и

значительно более «лиричных» авторов, каким, например, является И. С. Тургенев.

Среди самых крупных произведений, написанных по-русски в XX в., безусловно, романы М. А. Шолохова «Тихий Дон» (около 425 тыс. слов), «Понятая целина» (205 тысяч) и роман А. Н. Толстого «Хождение по мукам» (275,2). Многие известные авторы-романисты написали не более одного романа, превышающего 100 тысяч слов. Так, в романе М. А. Булгакова «Мастер и Маргарита» 112,4 тысячи слов, в «Чевенгуре» А. А. Платонова – 115, в «Даре» В. В. Набокова – 103. Остальные произведения этих авторов существенно меньше, особенно у Андрея Платонова. Условная шкала «эпичности»/«лиричности» не является сколько-нибудь строгой, но на ней четверка классиков расположилась так: Шолохов – Булгаков – Набоков – Платонов. И как бы мы ни объясняли склонность автора к крупным или, наоборот, малым формам, она так или иначе имеет личностный отпечаток.

2. Грамматико-статистические параметры текста

В этом и следующем разделе мы продолжим сравнивать четырех известных прозаиков XX в. и интерпретировать статистические данные, полученные на основании анализа 16 крупных текстов М. Булгакова («Белая гвардия», «Жизнь господина де Мольера», «Театральный роман», «Мастер и Маргарита»), В. Набокова («Король, дама, валет», «Защита Лужина», «Камера обскура», «Дар»), А. Платонова («Сокровенный человек», «Чевенгур», «Котлован», «Ювенильное море») и М. Шолохова («Тихий Дон», «Поднятая целина», «Судьба человека», «Они сражались за Родину»).

Можно привести следующие сведения о соотношении в этих текстах слов разных частей речи (точнее – знаменательных частей речи). Доля знаменательных слов у каждого автора колеблется от 54 до 60 %, в среднем 57–59 %. Однако обращают на себя внимание значимые различия в самом распределении слов разных частей речи.

В идиостиле Владимира Набокова проявляется явно повышенное количество признаков слов за счет уменьшения существительных и глаголов. Прилагательные составляют в его

произведениях более 16 % форм знаменательных частей речи, а если их объединить с наречиями, то в сумме получается 28 % (у других авторов другая цифра – 23 %. Это в целом подтверждает особенности употребления прилагательных В. Набоковым, часто намеренного, приводящего к компрессии художественного описания – см. об этом в работе [Мухин 1998].

М. Шолохов явно тяготеет к употреблению глаголов (почти 35 %), что обусловлено процессуальной динамичностью эпических произведений. Булгаков значительно чаще других употребляет в текстах числительные. Действительно, в его текстах многое подвергается счету (деньги, например). Часто встречается в романах Булгакова лексика, обозначающая время и его количественные параметры.

Теперь о длине предложений в исследуемых текстах: у М. Булгакова среднее количество слов в предложении – 11,3; у В. Набокова – 15,7; у А. Платонова – 14; у М. Шолохова – 9,7. При этом короче всего фразы в «Тихом Доне» – 9,2 слова; наиболее длинные зафиксированы в романе Набокова «Дар» – в среднем почти 22 слова. Можно было бы «привязать» эти сведения к разнице в сюжетах произведений. Действительно, длинные фразы встречаются там, где больше пространных рассуждений («Дар»), а короткие предложения – там, где, во-первых, быстрее разворачивается динамичное повествование и, во-вторых, больше диалогов. Рядом с «Тихим Доном» и «Поднятой целиной» здесь можно поставить «Белую гвардию» Булгакова или «Сокровенного человека» Платонова (около десяти слов в предложении).

Однако эта особенность имеет и явный идиостилевой оттенок. Если мы посмотрим на произведения Владимира Набокова, то в среднем в предложениях самых «динамичных» его романов «Король, дама, валет» и «Камера обскура» как минимум 12 слов. В итоге сюжетное влияние здесь оказывается не единственным фактором. Мы можем констатировать и авторские признаки: склонность В. Набокова и А. Платонова по-разному, но весьма пространно описывать предметы и ситуации и, наоборот, тяготение М. Булгакова и М. Шолохова к динамичным и даже отрывочным элементам романного повествования.

3. Лексика, авторский словарный запас

Как писал Гейр Хьетсо в связи с определением авторства «Тихого Дона», «богатство авторского словаря предпочтительно оценивать посредством коэффициента типичности знаков (КТЗ), т. е. измерением отношения числа различных форм (типов) слов к общему числу словоупотреблений (знаков) в тексте» [Хьетсо и др. 1989: 105]. Этот относительный коэффициент позволяет при оценке словарного запаса снять проблему влияния объема текста.

При этом с увеличением объема произведения это отношение падает: чем больше роман, тем относительно меньше уникальных слов приходится на целый текст. «По мере удлинения текста возможность обнаружить слова, ранее не употребленные писателем, естественно сокращается» [Там же: 106]. Так, у Булгакова в «Театральном романе» леммы составляют 25 % по отношению к словоформам, а в «Мастере и Маргарите» – 15 %; у Шолохова в рассказе «Судьба человека» – 42 %, а в «Тихом Доне» – 8,5 %. Для того, чтобы коэффициент типичности знаков стал объективным, нужно вычислить его среднее значение в сходных по объему частях текста.

Все исследуемые тексты были разбиты на фрагменты по 6 тыс. употреблений слов знаменательных частей речи (а не всех слов, как в исследовании скандинавского коллектива, возглавляемого Хьетсо). В каждом фрагменте вычислено количество лемм и выведено среднее соотношение КТЗ для всего произведения. Согласно полученным данным, КТЗ для слов знаменательных частей речи колеблется от 36 % («Котлован») до 53 % («Дар») слов на фрагмент, в основном около 44 %. По среднему КТЗ, т. е. условному лексикону, больше всего слов знаменательных частей речи у Михаила Шолохова (46,6 %), меньше всего – у Андрея Платонова (40 %). Примерно одинаковые значения КТЗ у Владимира Набокова (44,6 %) и Михаила Булгаков (44,3 %). Более подробно о методике подсчета см. в работе [Мухин 2010].

Конечно, из этого не следует, что автор с большим значением КТЗ лучше других или даже, например, имеет большую эрудицию. «То, с чем мы здесь сталкиваемся, – пишет Г. Хьетсо, – это прежде всего важная особенность стилистической манеры

писателя» [Там же: 114]. В чем же заключается эта важная особенность? В шолоховских текстах объем словаря, конечно, определяется большим количеством диалектных и просторечных слов и вариантов. Можно отметить лексическую «дотошность» Владимира Набокова, стремление к детализации, которую мы наблюдаем и у Набокова, и у Булгакова. И, безусловно, у Платонова к лексическим изыскам было достаточно спокойное отношение, что подтверждает идею Иосифа Бродского: Андрей Платонов не занимался «хитросплетениями сюжета, типографскими изысками и стилистическими кружевами» [Бродский 2001: 73]. Что это как не психолингвистическая характеристика творчества? В текстах писателя вроде так же много просторечных и узкоспециальных слов – однако совсем не так же много, как у М. Шолохова. В итоге платоновский лексикон оказывается по объему меньше, чем у других авторов.

Итак, мы рассмотрели одни из самых формальных признаков художественного текста: объем, распределение слов разных частей речи, среднюю длину предложения и условный словарный запас, понимаемый как отношение количества слов к количеству словоупотреблений. Чем формальнее признак, тем проще получить по нему недвусмысленные статистические данные. Мы видим, что эти данные свидетельствуют об идиостилевых особенностях текстов, и нам хочется видеть в этой интерпретации, пусть не абсолютно точной, психолингвистические черты. Естественно, судить о психологических особенностях самих авторов – реальных людей – по этим данным можно только косвенно.

ЛИТЕРАТУРА

Бабенко Л. Г. Филологический анализ текста. Основы теории, принципы и аспекты анализа. – М., Екатеринбург, 2004.

Баевский В. С. Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. – М., 2001.

Болотнова Н. С. Художественный текст в коммуникативном аспекте и комплексный анализ единиц лексического уровня. – Томск, 1992.

Бродский И. Послесловие к «Котловану» А. Платонова // Соч. Иосифа Бродского. – СПб., 2001. Т. 7. С. 72–74.

Казарин Ю. В. Поэтическое состояние языка (попытка осмысления). – Екатеринбург, 2002.

Лукин В. А. Художественный текст: основы лингвистической теории и элементы анализа. – М., 1999.

Мухин М. Ю. Лексическая статистика и концептуальная система автора: М. Булгаков, В. Набоков, А. Платонов, М. Шолохов. – Екатеринбург, 2010.

Мухин М. Ю. Синтагматическое напряжение в романах Владимира Набокова : дис. ... канд. филол. наук. – Екатеринбург, 1998.

Новиков Л. А. Художественный текст и его анализ. – М., 2007.

Пищальникова В. А. Проблема идиостиля. Психолингвистический аспект. – Барнаул, 1992.

Сулименко Н. Е. Текст и аспекты его лексического анализа. – М., 2009.

Хьетсо Г., Густавссон С. и др. Кто написал «Тихий Дон»? – М., 1989.

Чернухина И. Я. Основы контрастивной поэтики. – Воронеж, 1990.

Burrows J. F. Computers and the Study of Literature // Computers and Written Texts. – Oxford, 1992. Pp. 167–204.

©Мухин М. Ю., 2019